Model-Informed Approaches and Innovative Clinical Trial Design for Adeno-Associated Viral Vector-Based Gene Therapy Product Development: A White Paper

Amitava Mitra[1,*], Mariam A. Ahmed[2], Rajesh Krishna[3], Kefeng Sun[2], Francis D. Gibbons[4], Olivia Campagne[2], Noha Rayad[5], Youssef M. Roman[6], Salwa Albusaysi[7], Maria Burian[8] and Islam R. Younis[9]

The promise of viral vector-based gene therapy (GT) as a transformative paradigm for treating severely debilitating and life-threatening diseases is slowly coming to fruition with the recent approval of several drug products. However, they have a unique mechanism of action often necessitating a tortuous clinical development plan. Expertise in such complex therapeutic modality is still fairly limited in this emerging class of adeno-associated virus (AAV) vector-based gene therapies. Because of the irreversible mode of action and incomplete understanding of genotype–phenotype relationship and disease progression in rare diseases careful considerations should be given to GT product's benefit–risk profile. In particular, special attention needs to be paid to safe dose selection, reliable dose exposure response (including clinically relevant endpoints), or creative approaches in study design targeting small patient populations during clinical development. We believe that quantitative tools encompassed within model-informed drug development (MIDD) framework fits quite well in the development of such novel therapies, as they enable us to benefit from the totality of data approach in order to support dose selection as well as optimize clinical trial designs, end point selection, and patient enrichment. In this thought leadership paper, we provide our collective experiences, identify challenges, and suggest areas of improvement in applications of modeling and innovative trial design in development of AAV-based GT products and reflect on the challenges and opportunities for incorporating MIDD tools and more in rational development of these products.

Gene therapy (GT) is a therapeutic strategy that entails modification of defective genes in order to treat diseases.[1,2] GT works by either replacing a gene that is missing or dysfunctional or by turning off problem-causing genes. This therapeutic approach holds enormous promise for treating patients with genetic diseases, which is evident from the several product approvals in recent years. Of the GT platforms, the adeno-associated virus (AAV) vector-based platform is by far the most mature. Briefly, this platform works by introducing a transgene of interest into the host cell nucleus using a replication deficient AAV vector.[3] Currently, there are seven approved AAV-based GT products, as listed in **Table 1**.

As the prevalence of GT products in the development pipeline increases, there is a need to develop or to adapt clinical development, clinical pharmacology, drug metabolism, and pharmacokinetics (DMPK), and modeling and simulation tools to be able to adequately characterize and predict the performance of GT products in humans. Large, randomized, controlled clinical trials and full clinical pharmacology packages are often not available or feasible for GT products due to the recruiting challenges of patients with rare diseases,[4] poor understanding of pathophysiology and disease progression, and limited knowledge of clinically relevant end points to name a few. To this end, the US Food and Drug Administration (FDA)[5] and the European Medicines Agency (EMA)[6] have issued a series of draft and final guidance documents to guide clinical development. Additionally, several publications have provided in-depth review of clinical pharmacology[7] and DMPK[8] considerations in development of GT products.

There are two main facets to model-informed drug development (MIDD): (i) setting up dose and design considerations before initiating a clinical trial, and (ii) analyzing information streams and data once the clinical trials are completed to identify important covariates, developing dose response, and other relationships.

[1]Clinical Pharmacology, Kura Oncology, Boston, Massachusetts, USA; [2]Quantitative Clinical Pharmacology, Takeda Development Center Americas, Inc., Cambridge, Massachusetts, USA; [3]Integrated Drug Development, Certara USA, Inc., Princeton, New Jersey, USA; [4]Quantitative Solutions, Preclinical and Translational Sciences, Takeda Development Center Americas, Inc., Cambridge, Massachusetts, USA; [5]Clinical Pharmacology, Modeling and Simulation, Parexel International (MA) Corporation, Mississauga, Ontario, Canada; [6]Department of Pharmacotherapy and Outcomes Science, Virginia Commonwealth University School of Pharmacy, Richmond, Virginia, USA; [7]Department of Pharmaceutics, Faculty of Pharmacy, King Abdulaziz University, Jeddah, Saudi Arabia; [8]Translational Medicine Neuroscience and Gene Therapy, UCB Biopharma SRL, Braine-l'Alleud, Belgium; [9]Clinical Pharmacology Sciences, Gilead Science, Inc, Foster City, California, USA. *Correspondence: Amitava Mitra (amitra@kuraoncology.com)

Received March 31, 2023; accepted June 2, 2023. doi:10.1002/cpt.2972

Table 1 Approved AAV Vector-Based Gene Therapy Products

Gene therapy product	Generic name	Class	Sponsor	Date of approval	Indication
Glybera *(withdrawn from market)*	Alipogene tiparvovec	AAV gene therapy	UniQure Biopharma	2012	Familial lipoprotein lipase deficiency (LPLD)
Luxturna	Voretigene neparvovec-rzyl	AAV gene therapy	Spark Therapeutics	2017	Retinal dystrophy
Zolgenesma	Onasemnogene abeparvovec	AAV gene therapy	Novartis	2019	Spinal muscular atrophy
Hemgenix	Etranacogene dezaparvovec	AAV gene therapy	CSL Behring	2022	Hemophilia B
Roctavian	Valoctocogene roxaparvovec	AAV gene therapy	Biomarin	2022	Hemophilia A
Upstaza	Eladocagene exuparvovec	AAV gene therapy	PTC Therapeutics	2022	L-amino acid decarboxylase (AADC) deficiency
Elevidys	Delandistrogene moxeparvovec	AAV gene therapy	Sarepta Therapeutics	2023	Duchenne muscular dystrophy

AAV, adeno-associated virus.

Within the MIDD toolkit, there are several tools available to the data scientist, the type of analytic tool depending on the stage of development and appropriateness of the information.

Rational dose selection and dose optimization during GT clinical development is a field of intense research. Given the practical challenges of translating preclinical data to first-in-human (FIH) dose selection, and ethical challenges in developing GT products, which are more often developed for treatment of pediatric rare disease, MIDD tools are essential for appropriate dose selection and justification of these products. One of the first publications in this field by Tang *et al*.,[9] explored the utility of allometry in GT. Subsequently, a cross company perspective paper[10] emphasized the need for an integrated experimental and modeling approach early in development to identify dose response relationships across species, including the target patient population. To achieve this, it is critical to design thoughtful preclinical and clinical experiments to understand disposition of the vector and transgene, transduction efficiency, duration of expression, pharmacological effect, and safety signals. Recent publications from the FDA summarized the use of MIDD approaches in GT[11] and rare diseases,[12] in terms of optimizing dose regimen, supporting pediatric extrapolation, informing clinical trial design, and providing confirmatory evidence for effectiveness. These concepts and lessons learned can also be applied in GT product development.

In this whitepaper, we provide a comprehensive overview on the applications of several MIDD approaches in dose selection and innovative clinical trial strategies to develop AAV-based GT products. We attempt to suggest potential use of various tools for the benefit of future developers of AAV-based GTs.

MODEL-INFORMED APPROACHES FOR DOSE SELECTION AND OPTIMIZATION

Unlike other modalities, for which a range of doses are typically selected for clinical study, the complexities around AAV-based GT require more precision around dose selection. Often, the dose range is restricted by the safety and tolerability of the vectors in which the gene editing components are delivered. Moreover, because administration of AAV vectors may result in robust generation of anti-capsid antibodies, efforts to date have been constrained by only "one shot on goal" for such a therapeutic. Given the one-time therapeutic intervention with GT, it is even more critical for patient safety and product effectiveness that a totality of data approach is taken for dose selection and optimization, by integrating preclinical and clinical information through MIDD approaches.[13] Additional, FDA guidance on GT for rare diseases recommends that in serious or life-threatening diseases, study treatment should ideally start with a potentially therapeutic dose.[14] In this section, we summarize the use of several modeling approaches in dose selection and optimization of GT products.

Empirical and allometric scaling for efficacy

Traditional dose scaling approaches for FIH dose selection of GTs have been used.[7]

$$\text{Dose}_{\text{Human}} = \text{Dose}_{\text{Animal}} \times \text{Scaling Factor} \times \text{Activity Factor} \quad (1)$$

$$\text{Scaling Factor} = \left(\text{Body or organ metric}\right)_{\text{Human}} / \left(\text{Body or organ metric}\right)_{\text{Animal}} \quad (2)$$

This remains a continually evolving area, as it is uncertain which parameter of GT disposition would truly scale with body size. Summarized below are the current thinking on GT dose scaling based on body weight (BW) and organ volume, focusing on both efficacy and safety. Tang *et al*.[9] pioneered the allometric scaling concept with GT with a new parameter Gene Efficiency Factor (GEF) across species for calculating human dose for AAV-based factor IX (FIX) GT for hemophilia B. The GEF was defined as the ratio of the FIX protein synthesis rate (K_{syn}, mol/day) divided by the GT dose (vector genomes, vg/individual), and K_{syn} could be calculated as a product of the target concentration of the transgene protein and the latter's systemic clearance (CL; Eq. 3).

$$\text{GEF} = \frac{\text{CL}_{\text{Protein}} \times C_{\text{Protein}}}{vg\ \text{Dose}} \quad (3)$$

A roughly linear allometric relationship between log(GEF) and log(BW) was identified for three AAV-based gene therapies with FIX transgene with FIX level data from nonclinical species and humans. Based on this allometric scaling relationship, GEF roughly decreases with increased BW. The higher GEF observed in organisms with smaller body size may be attributed to their higher metabolic rates and, thus, energy consumption needed for DNA and protein synthesis.[9] The GEF vs. BW relationship can, therefore, be used to project human dose of AAV GT, provided that the transgene protein is secreted and measurable, and that data from at least two nonclinical species are available.

In follow-up to the GEF method, Aksenov *et al.*[10] re-analyzed data in Tang *et al.*, to demonstrate that different power laws for each FIX construct can be obtained, using a power regression model (FIX concentration = a ′ Dose^{b}) to describe dose–FIX concentration relationships for AAV-FIX vectors. However, a nonlinear dose–response relationship was found to exist in most nonclinical species and patients with hemophilia B for plasma FIX, and this allometric scaling may not accurately predict human dose for AAV GT. Zou[15,16] then compared the accuracy of FIH dose projection among current methods for gene therapies for hemophilia A and B, and found that the ranking was allometric scaling > dose–response normalization > direct vg/kg conversion for majority of the AAV vectors analyzed. For allometric scaling, the total amount of transgene product in blood circulation across species could be normalized to a species-invariant scale using an exponent of −0.25 for body weight ($\text{BW}^{-0.25}$). The GEF approach could then be refined as regression across species between log(GEF) and $\text{BW}^{-0.25}$. This method offered a more accurate human dose projection for the nine AAV vectors included in the dataset (in the absence of T-cell responses) and was an improvement over the original GEF method.

Empirical and allometric scaling based on safety

Safety-based dose projection for GT provides a ceiling for the vector dose to be administered during escalation and pivotal phases of clinical trials. In contrast to the multiple methods being proposed for efficacy-related dose scaling, the generally used safety-based dose scaling can be summarized as Eqs. 4 and 2:

$$\text{Dose}_{\text{Human}} = \text{Dose}_{\text{Animal}} \times \left(\text{Scaling factor}\right) \tag{4}$$

The total maximal tolerated dose (MTD) in animals (in total number of vg) from nonclinical safety and toxicity studies is adjusted by a morphologically based scaling factor to derive the upper limit of total dose (also in vg) in humans. The scaling factor in Eq. 2 is dependent upon the route of administration of the GT modality. For intravenous administration, the scaling factor is typically BW; therefore, the MTD in animals, when expressed in vg/kg, becomes the putative MTD (also in vg/kg) in humans, as seen in the cases of onasemnogene abeparvovec,[17] valoctocogene roxaparvovec,[18] and etranacogene dezaparvovec.[19] For direct injection into an organ/tissue with a clear physiological boundary, such as direct delivery into certain regions of the brain, the scaling factor is the human-to-animal ratio of the weight or volume of the tissue where the GT modality is delivered. Known examples include AMT-130 (direct infusion into the putamen and caudate nucleus)[20] and eladocagene exuparvovec (direct infusion into the putamen),[21] both of which used the human-to-nonhuman primate volumetric ratio on their respective target brain region(s) as the scaling factor to calculate the MTD.

In summary, as the most widely used method for GT dose projection, empirical and allometric scaling should be made based on both efficacy and safety observations of the modality. The method of scaling is dependent on the route of administration and the organ of tropism for the GT, as well as the measurability of the transgene product. Further refinement of the dose scaling approach should be explored, including inclusion of scaling factors identified in *in vitro–in vivo* correlation exercises.[22]

Translational models

Given the emerging state of both viral vector- and nonviral vector-based GTs, the translational science function in any enterprise should involve model-interconnectedness spanning discovery biology, vector engineering, therapeutic delivery, clinical sciences, data sciences, and clinical pharmacology functions. This allows for seamless integration of data, but also a data-driven approach to deliver functionality identification and optimization. Before delving into these models, we clarify certain terminologies specific to AAV GT[7,8]: (i) "pharmacokinetics (PKs)" of AAV is the *in vivo* disposition of both the capsid and vector genome, (ii) "biodistribution (BD)", is the *in vivo* distribution, persistence, and CL of AAV from the site of administration to tissues including bodily fluids, (iii) "shedding" is the excretion of the vector through excreta and secreted fluids, and (iv) "exposure" is the level of transgene mRNA and protein.

Exposure-related end points. Although the GEF[9,15] can be a useful approach to scaling across species at a gross level, it is more instructive to break down the efficiency of various aspects of transduction for comparison across serotypes and species. **Figure 1** illustrates these aspects of recombinant AAV-mediated GTs, with a focus on efficiency of delivery (i.e., what fraction of injected dose is detectable in the liver; **Figure 1a**) and transcription (i.e., how many transgene mRNAs are produced for each vector genome delivered; **Figure 1b**).

First, for vector delivery efficiency (**Figure 1a**), there can be significant variability within species for the same combination of serotype and transgene product—some of this is due to inter-individual variability (at the same dose), and some is due to the well-known nonlinearity of dose-exposure.[7] Among various AAV serotypes, they may have different tissue preferences, or "tropisms," for example, AAV5 and AAV6 do not deliver the vector to the liver as efficiently as other serotypes. Last, it is notable that median trends of vector delivery efficiency to the liver do not seem to differ significantly across species (i.e., mostly within threefold after adjustment by dose), although data are still rather limited.

Second, for transcription efficiency (**Figure 1b**), similar to vector DNA delivery efficiency, the somewhat sparser data showed significant variability exists between individuals dosed with the same modality. However, the most striking feature is that larger species (e.g., non-human primates and humans) show

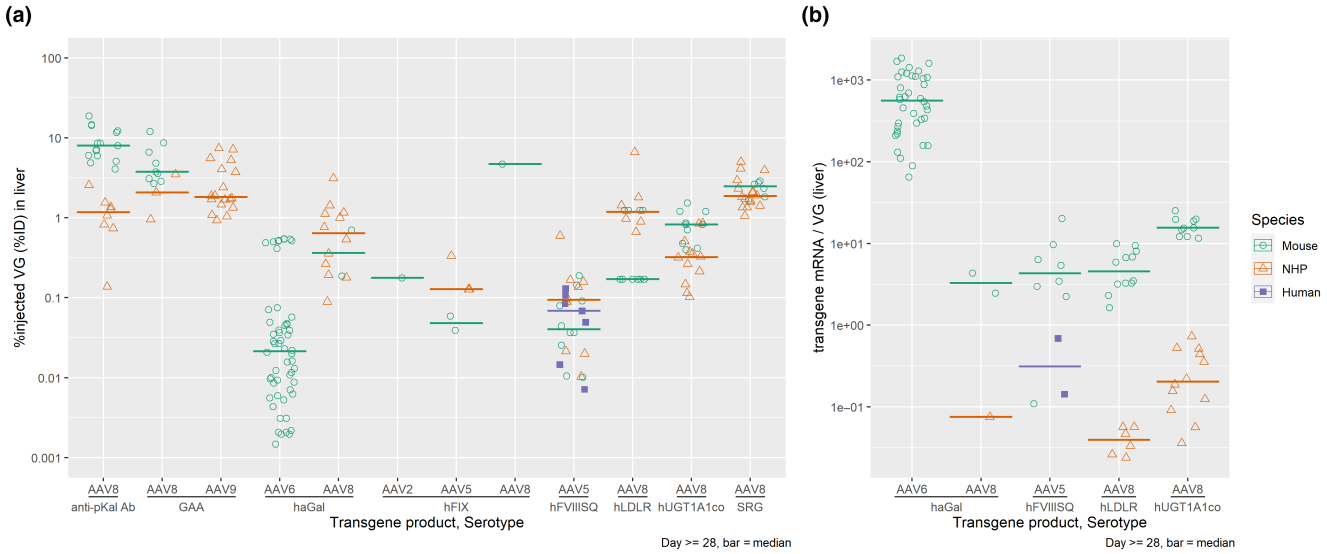

Figure 1 Delivery efficiency (**a**) and transcription efficiency (**b**) for AAV-mediated gene therapies. Shapes and colors indicate data for individuals from several species (○ = mouse, Δ = NHP (cynomolgus and rhesus monkeys), ■ = human), all data are for days ≥ 28 post-administration. Horizontal bars indicate group medians. AAV, adeno-associated virus; NHP, nonhuman primate.

greatly reduced transcription efficiency compared with rodent species. This can be attributed to systematic differences in cellular metabolism, perhaps driven by rate-limiting energy delivery to cells in larger species.[23,24] Where data are available, the drop-off in transcription efficiency appears to be approximately at least 20- to 30-fold. Whereas the stark differences in transcription efficiency in the liver have been demonstrated in multiple cases (**Figure 1**), regardless of the promoters used, much less is known on the similarity or potential difference in transcriptional regulation for GT modalities targeting the nervous system, muscles, and sensory organs. Limited data suggest that the drastic interspecies difference in transcription efficiency of exogenous vector DNA in the liver may not be present in the peripheral nervous system[25] or in skeletal muscles.[26] Furthermore, limited data also indicate that the overall rates of translation (from mRNA to intracellular polypeptide) and secretion (from intracellular transgene protein to extracellular) may be similar across species,[27] and that levels of host chaperones may contribute to individual variability in rates of secretion, and thus individual levels of the transgene protein.[28]

Pharmacodynamics-related end points. The widespread use of animal disease models in GT research offers crucial preclinical assessment of a candidate's potential in disease modification. Nevertheless, it is important to note that knockout/knock-in models differ from patients in terms of physiology, lifespan, target gene sequence, and disease onset and severity; no animal model fully represents a human disease in its entirety. Therefore, particular attention should be paid to non-similarities between animal models and patients when applying forward translation strategies. A mechanistic PK/pharmacodynamic (PD) or quantitative systems pharmacology (QSP) model may be able to incorporate the animal-human differences in physiology genetic background, biomarker levels, and disease progression.[7]

Taken together, the forward translation approach for GT should be informed by important mechanistic data that connects vector disposition and safety and efficacy. In particular, key determinants of transduction, transgene product PKs and PDs should be identified as early as possible in a GT project's lifecycle.

Mechanistic models

Mechanistic models, such as QSP and physiologically-based PK (PBPK) models, integrate PK, physiology, and biological pathways and have been developed for small molecules, therapeutic proteins, and RNAs.[29,30] The same principle in development, verification/validation, and application of mechanistic models readily applies to GT modalities. As of 2023, the scope of such model for GTs generally covers the kinetics of one or more of the following aspects[31]: (i) biodistribution of the vector; (ii) transcription, and translation of the transgene product; (iii) kinetics of the transgene product; and (iv) PDs and pharmacology of the transgene product.

Modeling vector biodistribution. Various QSP/PBPK models describe the kinetics of the viral vector from immediately after dosing to the eventual formation of exogenous DNA inside the host cell nucleus.[32,33] These models are typically multiscale (i.e., being capable of characterizing vector concentrations both at a systemic or organ-level and in (sub)cellular compartments). Platform-like biodistribution models are generally derived from rich time course data of vector concentrations gathered from rodent experiments and then combined with well-known organ physiology models.[34,35] Re-calibration of key parameters, such as rates of target tissue uptake, degradation, and uncoating of the vector, is necessary when adapting such platforms for vectors of different serotypes and for different species. In addition, potential models of immune response to the vector[36,37] can be incorporated for more rational exploration of loss of transduction.

Modeling transcription, translation, and secretion of the transgene protein. Well-defined minimal quantitative systems biology models for the formation, transport, and degradation of mRNA and protein[38] can be adapted to QSP models for GT and linked to the vector DNA kinetic model. Calibration of the formation rates is generally necessary, as these vary intrinsically among transgene sequences. Further calibration with cross-species mRNA and transgene product concentration data is often required due to species difference in metabolism (see Translational Approaches section).

Modeling PKs and PDs of the transgene product. A disease-specific PK/PD model for the transgene product, if available, may be connected to the QSP model for GT vector biodistribution, transcription, and translation, to result in an integrated QSP model that begins with the vector dose and ends with the projected pharmacological effects, including clinical assessment endpoints (**Figure 2**). The prerequisite for such a model, other than those already listed, also includes the availability of quantitative assays for both the transgene product and its PD biomarker(s)/response measurement. It is necessary to consider the potential species differences in the PKs of the transgene protein, as well as in disease biomarker levels.

QSP models of GT can directly inform designs of nonclinical and clinical studies. Early platform-like models can be used for lead optimization and for selection of timepoints and end points of nonclinical experiments, whereas a mature QSP model that incorporates disease modeling could inform selection of clinical doses, planning schedule of assessments, and proof-of-concept evaluations. The QSP model for GT can also be particularly useful to benchmark the GT against other modalities' PK and PD data in the same disease area, such as small molecule modulators and enzyme/factor replacement therapies. Challenges in development and usage of these QSP models include time and resources required for building a *de novo* platform model, as well as sparseness of vector biodistribution data in higher species required for calibration. However, progress has been made in PBPK modeling of biodistribution as demonstrated by Sun *et al.*[32] Thus, existing platform models should be used in early stages of a project; other techniques, such as meta-analysis of historical biodistribution data, may be used to enhance the data richness in higher species.

Dose/exposure-response models

It is important to characterize the relationship between dose and eventual transgene expression as well as that between dose and clinical efficacy to achieve optimal efficacy of drugs including GT medications. Dose–response analysis can be utilized to optimize dosing of GT products.

Onasemnogene abeparvovec[39] is an AAV vector-based GT indicated for the treatment of pediatric patients less than 2 years of age with spinal muscular atrophy. Dose–response relationship provided supportive evidence of efficacy. Two open-label, single-arm clinical trials were used to support approval. One trial enrolled 15 participants, 3 in a low-dose group and 12 in a high-dose group.

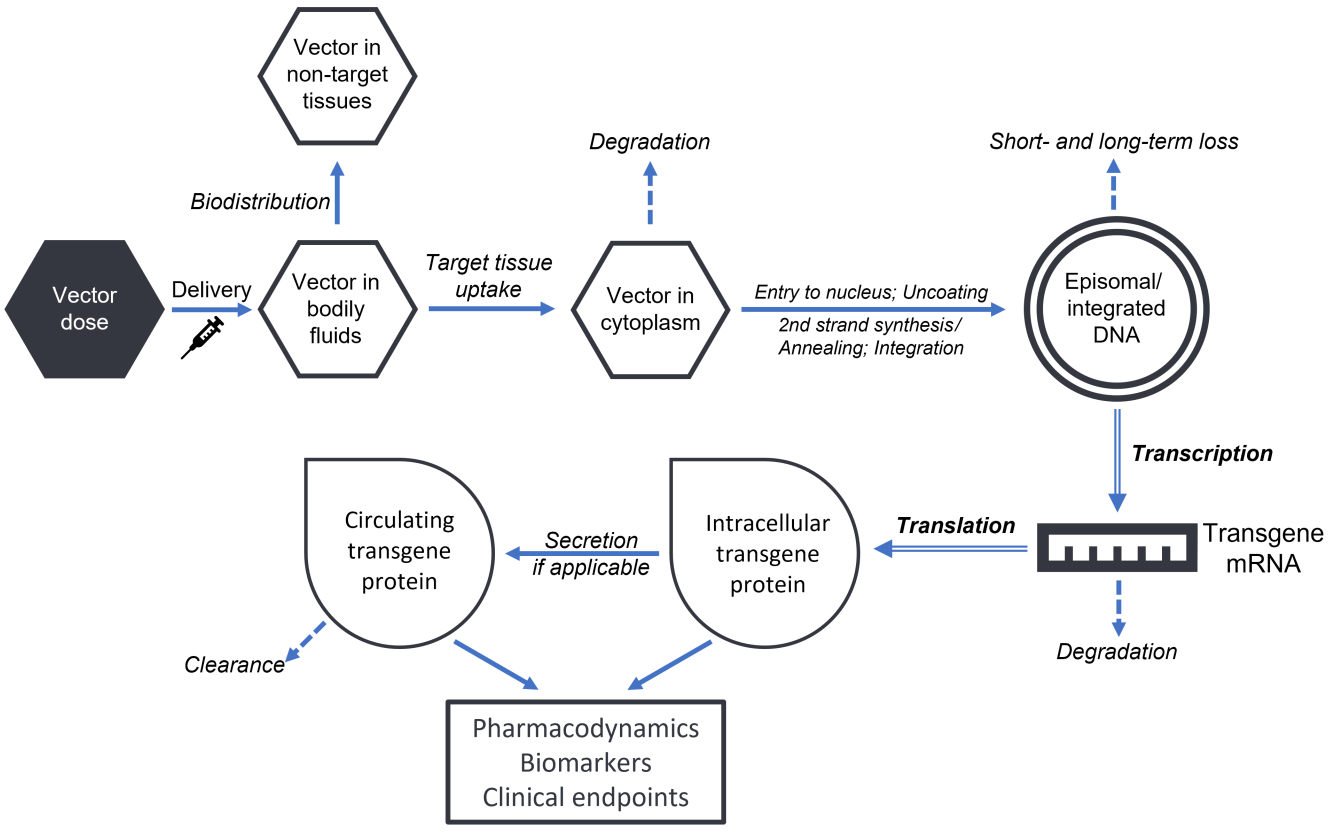

Figure 2 A quantitative systems pharmacology (QSP) framework for gene therapy modalities.

One participant in the low-dose group had to receive permanent ventilation (an indication of therapy failure) and all 12 patients in the high-dose group were alive without permanent ventilation after 24 months of drug administration, which reflects the success of therapy.

Dose–response analysis has also been used to characterize the effect of GT in preclinical studies.[40] The authors conducted detailed analysis of transduction levels throughout the brain, brainstem, and spinal cord of nonhuman primates and demonstrated that intrathecal cerebrospinal fluid delivery of the GT combined with tilting (Trendelenburg position) leads to widespread transduction in the brain and spinal cord of nonhuman primates. Additionally, 30 times lower intrathecal doses compared with i.v. injections led to transduction of up to 55–80% motor neurons in all regions of the spinal cord in nonhuman primates. Such studies along with dose–response analysis offers critical insights in vector distribution depending on route of administration and its correlation with transgene expression.

The use of dose response analysis in preclinical species to predict FIH dose and human dose–response relationship has also been proposed.[15] Dose–response for AAV-based hemophilia GT product across three preclinical species were normalized to a species invariant scale and subsequently used for FIH dose prediction. However, this approach was demonstrated to be less accurate in FIH dose prediction as compared with allometry using an exponent of 0.25 for GEF.

ADDITIONAL ASPECTS OF CLINICAL TRIAL DESIGN INFLUENCED BY MIDD

GT has no doubt come of age with several approved products in the past few years. However, challenges with GT clinical trials remain,[41] including but not limited to selecting appropriate patient population, small patient numbers (rare diseases), adequate understanding of the natural history (NH) of the disease to appropriately design a study, assessing treatment benefit in a reasonable time duration, and clear understanding of the pathophysiology and disease progression to quantify effectiveness and safety. To alleviate some of these challenges, several novel approaches have been adopted by GT trial sponsors. In this section, some of these strategies are reviewed based on the current GT clinical trials. Some key characteristics of the clinical trials for the approved AAV vector-based gene therapies in the United States and the European Union are summarized in **Table 2**.

Synthetic and external controls

Although external control arms (ECAs) have not been applied for AAV trials, they do represent an opportunity to streamline clinical development. The concept of ECA has received recent widespread recognition in pediatric and adult rare disease clinical trials.[42] An ECA involves the use of patient controls that are not a part of the same clinical trial, affording the value of borrowing information in cases where trials are constrained by small patient populations. Regulatory health authorities have in-principle supported the use of such ECAs in clinical trial and program approval decisions, including a very recent guidance from the FDA.[43] However, there are a plethora of control groups that have been variably used in statistics, including but not limited to synthetic controls, NH controls, historical controls, and external comparators. All of these variations are subsets of ECAs. Synthetic controls are derived from the outputs of MIDD, namely modeling and simulation that encompasses both pooling as well as aggregate analyses, such as meta-analyses and are the focus of this section.

A pertinent example lies in the use of synthetic control derived model-based meta-analysis to support the accelerated approval of blinatumomab in adult patients with relapsed/refractory acute lymphoblastic leukemia based on a single-arm trial with < 200 patients.[44] Modeling and simulation were used to evaluate the effect of blinatumomab compared with available therapies for proportion of complete remission, duration of complete remission, and overall survival using meta-analysis models and clinical trial simulations.

Such use of synthetic controls can be readily applied in several GT development programs. Diseases such as hemophilia lend themselves to synthetic controls very well because of pressing logistical and ethical considerations that preclude the use of randomized clinical trials in this disease population. The lack of a common comparator control among published trials makes such comparisons challenging to better delineate safety and efficacy of GT as compared with the standard of care. Real-world evidence (RWE) offers significant benefits to developed synthetic controls. The availability of registry studies in hemophilia, such as the World Federation of Hemophilia GT registry can facilitate the development of such controls.[45] Garrison *et al*.[46] further recommends that GT trials should wherever possible collect longitudinal data prior to and after therapy treatment and such lead-in self-controlled trials can yield a data source that may be more acceptable to regulators vs. the use of physician and patient collected retrospective data.

It is imperative that a collaboration between the statistical scientists and pharmacometricians exists to ensure that the best methodology of external controls, in particular the "dynamic borrowing" is applied when trials with small sample sizes are planned.

Application of natural history data in GT trials

The NH of a disease is defined as the course a disease takes from its onset until either the disease's resolution or the individual's death, in the absence of intervention to the disease.[47,48] An NH study is a preplanned observational or retrospective study intended to track the course of the disease. The purpose of such a study is to follow a group of people over time who either have or are at risk of developing a specific disease, with the goal to identify variables (e.g., genetics and demographics) that might correlate with the disease's development, and progression. Two FDA guidances clearly mention the need for the control and treatment populations to be well matched in terms of demographics, disease state, etc., for the NH data to form the basis of control for approval of the product, among other requirements.[14,48]

In the development of GT products for rare diseases, NH data can serve as a comparator arm to assess clinical outcomes (safety and efficacy) of interventional therapy, when it may be impractical and/or unethical to randomize patients to placebo. Additionally, NH data can provide important insights into—(i) *a priori* identification of patient population which can expedite recruitment into the interventional trial, (ii) establishment of relevant clinical

Table 2 Key Characteristics of the Clinical Trials for the Approved AAV Vector-Based Gene Therapy in the United States and the European Union

Brand	Primary efficacy end point	Is the primary end point surrogate biomarker, intermediate, novel, and/or clinical outcome?	Key secondary and PD end point	Additional confirmatory trials for efficacy	Sample size in pivotal trial	Enrichment strategies
Glybera *(withdrawn from market)*	Initial: reduction in fasting plasma triglyceride (median of baseline vs. median of week 3–12 post-treatment) ≥40%(a)[a]	Unvalidated surrogate	Reductions of fasting median chylomicrons and /or chylomicrons-triglyceride ratio 12 weeks post-treatment	Yes	27	Predictive and prognostic
Luxturna	1-year change in MLMT performance, measuring functional vision at specified light levels	Intermediate/novel	FST testing, visual acuity, and visual fields	No	31	Predictive
Zolgensma	Co-primary end points: survival at 14 months of age and the proportion of patients who achieved functional independent sitting for ≥ 30 seconds at the 18 months of age study visit	Long- and intermediate-term clinical outcome	Proportion of subjects independent of ventilatory support at age 18 months and proportion of subjects maintaining the ability to thrive at age 18 months	No	23	Predictive and prognostic
Hemgenix	ABR during months 7 to 18 (52 weeks) compared with ABR during the lead-in period (months 0 to 6)	Intermediate	Factor IX was assessed monthly in the first year then every 6 months	No	54	Predictive
Roctavian	Change from baseline in FVIII activity at 49 to 52 weeks after infusion	Surrogate	Change of the annualized utilization of exogenous FVIII replacement therapy and in the annualized number of bleeding episodes requiring exogenous FVIII replacement treatment during week 5 to the last visit	Yes	134 (ITT population)	Predictive and prognostic
Upstaza	Achievement of motor milestones as assessed by PDMS-2 at the 2-year timepoint	Intermediate	PDMS-2, AIMS, Bayley-III total subscale scores and neurological exam findings, change from baseline in CSF neurotransmitter metabolites	No	28	Predictive
Elevidys	Micro-dystrophin at 12 weeks (surrogate end-point used for accelerated approval)	Surrogate	Time to rise from floor at 48 weeks	Yes (confirmatory trial on-going)	41 (126 for confirmatory trial)	Predictive

AAV, adeno-associated virus; ABR, Annualized Bleeding Rate; AIMS, Alberta Infant Motor Scale; CSF, cerebrospinal fluid; FST, Full-field light Sensitivity Threshold; ITT, intention to treat; MLMT, multi-luminance mobility test; PD, pharmacodynamic; PDMS-2, Peabody Developmental Motor Scale, second edition.

[a]At first, the main measure of efficacy was based on a reduction in blood triglyceride levels. However, this was later changed to postprandial chylomicrons, as this biomarker was thought to more specifically address the pharmacodynamic effect of Glybera, whereas the effect on triglyceride levels was only short-lived.

end points, surrogate end points, and biomarkers, (iii) identifying patient reported outcomes and quality of life measures, and (iv) development of RWE to support any postmarketing requirement.

The expanding role of NH studies in drug development have been demonstrated in several rare diseases, such as spinal muscular atrophy,[49] RPE65-associated inherited retinal dystrophy,[50] Leber hereditary optic neuropathy,[51] RPGR associated retinopathy,[52] Rho-associated retinitis pigmentosa,[53] Duchenne muscular dystrophy (DMD),[54] and infantile neuronal ceroid lipofuscinosis type 2.[55] These NH data will provide clinicians, drug developers, and patients with more knowledge about the condition and inform the design and interpretation of interventional trials.

In addition to supporting clinical development, NH data are increasingly being used in regulatory reviews for GT product approval. In the case of Zolgensma (onasemnogene abeparvovec), efficacy was evaluated in open label single-arm clinical trials.[56,57] The evidence of clinical benefit, which formed the basis of approval of onasemnogene abeparvovec, came from NH data as referred to in the product label—"Comparison of the results of the ongoing clinical trial to available natural history data of patients with infantile onset SMA provides primary evidence of the effectiveness of ZOLGENSMA."[17]

In the case of Luxturna (Voretigene neparvovec), the FDA encouraged Spark Therapeutics to conduct an NH study in patients with biallelic *RPE65* mutation-associated retinal dystrophy because NH data could be useful in interpreting safety and efficacy data generated from the interventional trial.[41] An NH study was conducted in parallel to the phase III trial. The NH study involved a retrospective chart review of patients who had a genetically confirmed diagnosis of autosomal-recessive mutations in the RPE65 gene and at least two office visits prior to retinal surgery or enrollment in an interventional study. A database of individuals who met the inclusion criteria, was developed from which curves describing the loss of individuals' visual field and visual acuity over time were constructed.[50] In addition to having utility in assessing clinical outcome, these NH data might also provide insights into the optimal timing of treatment.

Statistical and/or other quantitative modeling of NH data has become of particular relevance in GT clinical development, as these model-based approaches can increase much needed confidence in these data, as compared with retrospective patient chart reviews. The information gathered from modeling of NH data can be used to justify end point selection, change of disease trajectory over time to guide interventional strategy, correlation with patient functional assessment, and identification of patient population. Such analysis will have significant impact on a GT program not only during clinical development but postapproval to convince payers to reimburse these expensive one-time therapies.

End point selection and trial duration

When designing clinical trials with GT, end point selection and clinical trial duration are key considerations not only to enable benefit risk assessment by regulators but to ensure adequate evidence to inform health technology assessment (HTA) bodies decision making. The use of biomarker or intermediate surrogate clinical outcome is ethically preferable, especially when clinical events are rare/delayed in slowly progressive diseases or when there is a high unmet need. It is also practically preferable because the relatively short-term assessment helps to avoid noncompliance and missing data, increasing efficiency and reliability of the study. However, this paradigm raises uncertainties especially regarding the clinical meaningfulness and the durability of the assessed response as well as the sufficiency of the safety database. Carvalho *et al*.[58] analyzed and compared the major objections reported in the marketing authorization application assessment for approved advanced therapy medicinal products (ATMPs; $n = 3$) and non-approved ATMPs ($n = 4$). The most frequent objections for GT medicinal products in terms of clinical efficacy were lack of or insufficient demonstration of efficacy, the change or use of novel and non-validated primary end points, and efficacy claims based on non-prespecified *post hoc* analysis. Regarding safety, the most common objections were the limited safety database and the risks associated with immunogenicity.

For the seven approved AAV-based GTs in the United States and/or the European Union, **Table 2** summarizes information relevant to the primary efficacy end point, such as the timing of the assessment for the primary analysis, whether it is a surrogate end point, intermediate vs. long-term clinical outcome, and whether confirmatory clinical efficacy studies were required postapproval. Aside from onasemnogene abeparvovec,[56,57] where survival was assessed as a co-primary end point, all other approvals were based on surrogate biomarker or intermediate clinical outcome. For Roctavian (valoctocogene roxaparvovec),[59] the sponsor relied on Factor VIII (FIIV) activity as a surrogate biomarker. Based on the limited efficacy data and the likelihood of submitting a more comprehensive data postapproval, the European Commission granted a conditional marketing authorization. The postmarketing long-term efficacy studies are currently undertaken to evaluate the long-term effect of a single dose administration of valoctocogene roxaparvovec on bleeding profile, quality of life, and durability of FVIII activity (projected up to 15 years).[60] On the other hand, early interaction with regulators has shifted the sponsor intention from using Factor IX activity as the primary end point to assess the efficacy of Hemgenix (etranacogene dezaparvovec)[61] to the more acceptable Annualized Bleeding Rate as an intermediate clinical outcome for hemophilia, thus enabling full approval. For voretigene neparvovec-rzyl,[62] the sponsor relied on a novel intermediate clinical outcome to ensure high power of the small trials.

For pharmacometricians, the small sample size and the limited duration of the follow-up pose certain challenges and specific considerations when analyzing and interpreting the data. Design optimization utilizing statistical theory of optimal designs can guide sample collections to maximize the amount of information in the experiment for the given objectives and resource constraints. To our knowledge, this technique has not been utilized in GT yet,[63] although the concept from other modalities is readily applicable to GTs. For example, as mentioned in the section of QSP modeling, a mature QSP model that integrates the PK/PD data from earlier clinical trials can be used to optimize the schedule of assessments for later phase clinical trials. For earlier phase trials, optimal sampling that relies on nonclinical data can be achieved using statistical techniques that incorporate parameter or model uncertainty.[64]

It is critical to understand the correlation between the surrogate/short-term clinical outcome and intermediate or long-term clinical outcome from early phase clinical trials to help design later phase clinical trials. Given the small sample size, pharmacometricians may analyze the data according to the extreme value theorem, such as block maxima analysis and peak over threshold analysis.[65] In addition, model-based meta-analysis can be applied to incorporate all the available longitudinal data available from early phase clinical trials or for other modalities to understand the relationship between a biomarker or short-term clinical outcome and intermediate or long-term clinical outcome.[66] Moreover, NH and real-world data (RWD) can be incorporated within a Bayesian-based clinical trial simulation to predict long-term clinical outcomes based on earlier time data.[67]

Disease progression and extrapolation of durability of response models can inform discussions with HTA bodies on the cost-effectiveness of GTs based on the predictions of the durability of response. For example, Shah *et al*.[68] used both Bayesian and frequentist linear mixed models to predict the FIX activity level up to 25.5 years post-etranacogene dezaparvovec infusion at both the individual and population level. FIX activity levels < 2% were assumed to correlate with a severe bleeding phenotype needing regular prophylactic treatment with FIX replacement products. Using a Bayesian linear mixed model, it was predicted that > 80% of future patients receiving etranacogene dezaparvovec would be free from prophylactic FIX replacement products for more than 25 years post-infusion. Although the model is informed only by 2 to 3 years' data from phase II and III clinical trials, the authors assert that there is no evidence of waning FIX level in patients with hemophilia B receiving GT based on an 8-year published cohort with another GT in patients with hemophilia B.[69] Cook *et al*.[70] used RWD from patients with hemophilia A receiving prophylactic FVIII and the publicly available clinical trial data for valoctocogene roxaparvovec (up to 3 years) to assess the cost-effectiveness of GT. This was achieved by modeling the long-term disease progression and the durability of the response by incorporating an initial treatment effect (max FVIII) and treatment waning over time, which are used more widely to determine gene treatment durability. When considering modeling the durability of response, given the uncertainties around the long-term GT use, pharmacometricians should consider evaluation of scenarios that consider different duration of effects based on varying assumptions on the time horizon. For example, for hemophilia GT products, a time horizon of 10 years of efficacy, ±2, 3, or 4 years has been suggested given the availability of long-term clinical trials with up to 8 years of follow-up.[71] Modeling efforts should be expanded to capture different clinical outcomes as well as the available PK data in a validated fashion to pressure-test the validity of the assumptions of the durability of the response.

Accumulation of knowledge has enabled some understanding of the exposure-safety relationship for viral-based GT. For example, it has been shown that for intravenously administered AAVs, there are clear positive relationships between the AAV capsid dose and both the incidence and severity of treatment-emergent adverse events. In addition, very high levels of transgene product could result in excess pharmacological activity.[7] Predictive or mechanistic safety evaluations can be combined with Bayesian statistical approaches to analyze data from the confirmatory studies.[72] Of note, safety data generated prior to marketing authorization is limited and it is a typical regulatory requirement to require long-term follow-up safety data (e.g., up to 15 years). To comply with the requirements for long-term follow-up, product registries are typically utilized, thus creating an opportunity for collecting RWD that can be utilized to inform future modeling of long-term safety and durability of response.

Enrichment in GT clinical trials

Clinical trials addressing rare genetic diseases among special populations are compounded by the heterogeneity of the disease phenotypes, and in several cases, absences of genotype–phenotype correlations. Population enrichment strategies in drug development programs can increase the probability of detecting a treatment effect.[73,74] Enrichment is the prospective use of any patient characteristic to select a study population in which detection of a drug effect (if one is in fact present) is more likely than it would be in an unselected population.[75] Generally, enrichment designs offer a great potential for increasing the power of studies to detect a real effect of a treatment and the likelihood of conventional drug development success. A systematic analysis of pediatric drug development programs submitted to the FDA between 2012 and 2016 highlighted an association between the number of enrichment strategies used and the success rate of drug development clinical trials among pediatrics.[74] Among 112 efficacy studies submitted to the FDA, prognostic strategies were the most frequently used strategy (41.5%). Additionally, a large majority of studies (76.8%) used at least one enrichment strategy. Of those, 66.3% used multiple enrichment strategies. In trials that used multiple enrichment strategies, the success rate was 87.5% when three strategies were used together compared with 81.4% for the use of any single enrichment strategy. Of note, the lowest success rate was 65.4% when no enrichment strategy was used. Although enrichment strategies could provide a great promise for drug development approval, they also introduce some limitations to generate RWE. In addition, the variability in measuring a biomarker to dichotomize the response to therapy could disadvantage the marker-negative population. Occasionally, using a specific biomarker, especially in predictive enrichment, does not accurately characterize the populations to responders or nonresponders. To this extent, the inclusion of some marker-negative patients is encouraged in most trials unless strong pathophysiological or mechanistic rationale exists. Therefore, sponsors are strongly encouraged to engage with regulatory agencies for guidance early in the development program.

Strategies to use population enrichment may include practical, prognostic, and predictive enrichments.[73] **Table 2** summarizes the enrichment strategies for recently approved AAV-based GT products. Practical enrichment strategies aim to decrease intrapatient and interpatient variability by selecting patients who are more likely to respond to the treatment given their diagnosis and adherence to the study protocol. Prognostic enrichment designs aim to increase the proportion of patients likely to have a particularly disease-associated event or significant worsening in the disease. To this extent, prognostic enrichment strategies aim to identify

patients with historical characteristics and the event of interest to show a risk reduction. This strategy is commonly used when the treatment is intended to delay the progression of a particular disease, such as in patients with multiple sclerosis with prespecified magnetic resonance imaging or patients with prostate cancer with high prostate-specific antigen. Predictive enrichment is an approach that aims to enroll participants with biomarkers that may indicate an increased chance of treatment response. This strategy is intended to increase the early efficiency and feasibility of clinical studies and enhance the benefit–risk relationship for patients in the enriched subset compared with the overall population. Examples of predictive enrichment strategies include optimal level of response to trastuzumab among patients with breast cancer expressing human epidermal growth factors receptor 2 or the response to ivacaftor in patients with cystic fibrosis with select mutations in the *CFTR* gene.

Implementing enrichment strategies echoes the growing interest in personalized or precision medicine. The ability to tailor the treatment to those who will respond while balancing the ethical obligations when marker-negative population will not respond therapy is of paramount importance in rare diseases. For example, a predictive enrichment strategy based on genomic measures can have a significant impact on the probability of treatment success among rare diseases linked to gene variants. This enrichment strategy was used during the development of voretigene neparvovec-rzyl. The safety and efficacy of voretigene neparvovec-rzyl was established in a clinical development program with a total of 41 patients between the ages of 4 and 44 years. All participants had confirmed biallelic RPE65 mutations, which are associated with retinal dystrophy.[76,77]

A prognostic enrichment has been demonstrated in valoctocogene roxaparvovec clinical trial. The phase III trial was a single-arm, open label study which enrolled 134 adult men with severe hemophilia A who were on standard prophylactic replacement therapy. All participants had severe hemophilia A at baseline, defined as less than or equal to 1 IU/dL of FVIII activity. Results of the study revealed that valoctocogene roxaparvovec was effective at increasing the level of FVIII activity and that this increase was sustained for a minimum of 2 years.[18]

Applications of MIDD approaches in the life cycle of drug development can optimize the inclusion and exclusion criteria and treatment enrichment.[78] For example, MIDD approaches, such as disease progression models, have been utilized to support the use of genetic mutations for patient enrichment among men diagnosed with DMD to account for phenotypic variability and time to loss-of-ambulation onset.[47,79] Modeling changes in DMD biomarkers across muscle phenotypes can be used to detect and monitor the therapeutic effects of different treatment modalities on disease and provide prognostic information on functional outcomes.[80] In addition, disease model progression driven by machine learning (ML) could optimally characterize distinct disease states and the probabilities of progressing through these states which may improve trial design and participant selection. For example, modeling the disease progression through the integration of biomarkers for disease severity or clinical outcome measures, could be used to identify a likely responder population and thus assist with selection of patients more likely to respond to the treatment. In addition, certain interventions aimed to improve or slow down disease progression may be more likely to demonstrate their effects at a later stage of the disease. Therefore, selection of patients with more rapid disease progression may increase the probability of showing an effect. Collectively, enrichment strategies informed by MIDD can provide novel trial designs with smaller sample sizes for studies that present challenges with patient recruitment due to low prevalence and heterogeneity of the genetic diseases, leveraging the findings across different drug development programs in the same disease population. MIDD approaches can provide substantial or confirmatory evidence to support study population enrichment by leveraging the findings across different drug development programs in the same disease population. However, MIDD approaches for enrichment in GT trials still warrant further research, because in theory, GT should be able to correct all types of mutations in the gene of interest. Hence, genetic mutations associated with more severe phenotypes may not be a predictive factor.

POST APPROVAL AND REAL-WORLD EVIDENCE

For all AAV-based gene therapies, lack of viable re-administration strategy and established treatment paradigms in genetic diseases as well as single dose nature raise concerns regarding long-term durability of response as well as safety. Thus, regulatory authorities require long-term postmarketing studies.

In the development of GT products, RWD and RWE can be sources of information for retrospective or prospective NH studies and provide evidence of patient benefit postapproval. However, it is important to clarify the definitions of RWD and RWE and avoid using them interchangeably. The FDA in its *Framework for FDA's Real-World Evidence Program*[81] defined RWD as "data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources" and RWE as "clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD." Similarly, the EMA defines RWD as "routinely collected data relating to a patient's health status or the delivery of health care from a variety of sources other than traditional clinical trials" and RWE as "information derived from an analysis of RWD."[82]

Randomized controlled trials (RCTs) are still considered the "gold standard" for clinical development primarily because they are less prone to bias and provide a rigorous investigation of the cause-and-effect relationship between treatment and outcome. However, in the era of advanced therapies, such as AAV-based GT, and increasing focus on rare diseases, conducting RCTs may not always be feasible or ethical. In these instances, single-arm interventional trials with supportive RWE-based external control data, such as from NH studies, provide a viable path for drug development and approval.

RWE can provide critical insights into the long-term safety and maintaining effectiveness of the therapy in a real-world setting.[83] Particularly in the case of GT for rare diseases, which are typically conditionally approved with less clinical data than would be normally required due to a variety of reasons, including limited patient pool and severity of the disease, RWD can bolster efficacy and safety information postapproval.

In the case of voretigene neparvovec-rzyl, a postapproval long-term follow-up study has been ongoing to generate evidence of safety and effectiveness of the treatment.[84] Data presented at the 2022 ARVO annual meeting[85] demonstrated that effectiveness of voretigene neparvovec-rzyl was consistent with previous clinical trial findings in terms of improvement of visual function up to 2 years post-treatment. This ongoing study also identified a new adverse drug reaction (chorioretinal atrophy) which has not been associated with loss of visual function. Similarly, for onasemnogene abeparvovec a post-treatment follow-up study for up to 17 months demonstrated that all patients showed improved motor function and there was no case of mortality or requirement for permanent ventilatory support.[86]

In a first by any health agency, the German health agency, has mandated the sponsor to collect RWE for onasemnogene abeparvovec to demonstrate long term clinical benefit.[87] Taking a step further, the agency has specified that the sponsor should carry out a registry study directly comparing onasemnogene abeparvovec with Spinraza (Biogen's SMA treatment product). The expectation is that this RWE would provide further evidence of long-term clinical benefit and might support reimbursement for the patients needing the treatment. It should not come as a surprise if this type of post launch RWE generation becomes an expectation from health agencies and payers, because GT clinical trials are typically smaller in size and of shorter duration, as compared with RCTs. Therefore, whereas the initial data pool may be sufficient for conditional regulatory approval, it lacks sufficient long-term safety and efficacy data needed for full regulatory approval or reimbursement. These shortcomings of GT trials were summarized in a 2019 publication on voretigene neparvovec-rzyl, where the author concluded that additional review of the clinical studies revealed that the drug might not restore normal vision, vision improvement might not persist long-term, and patients experience vision loss.[88] Additionally, convincing the payers to provide reimbursement is also of strategic importance because that facilitates access to these expensive one-time therapies to the patients in need.[89]

Thus, it is imperative that developers of GT products have proactive interactions with regulatory agencies and HTA agencies to develop RWE strategies to satisfy both full approval and reimbursement needs. Such proactive efforts are critical for postapproval success of the product. Failing to do so can lead to commercial failure of the product, as was exemplified by Glybera (alipogene tiparvovec), which had to be withdrawn from the market after 2 years due to the costs of postmarketing requirements and extremely limited use.[90] MIDD in combination with RWD has the potential to revolutionize development of GT products for rare diseases by enhancing understanding of disease pathophysiology and guiding efficient clinical trial design. Indeed, several model-based and statistical methodologies have been developed and proposed for analysis of RWD[47] for understanding disease progression, prediction of treatment effect, clinical trial design, external control synthesis, and evidence generation for long-term treatment effects.

FUTURE DIRECTIONS AND CLOSING REMARKS

In recent years, the crucial role of MIDD in development of novel therapeutics such as siRNA[30] and cell therapy[91] has come to the forefront. Model-based approaches are poised to become the cornerstone of drug development in rare diseases, and specifically in GT due to the practical and ethical challenges in development of GT products for rare diseases, where large, randomized, controlled clinical trials and generation of full clinical pharmacology packages are often not feasible. Hence, MIDD approaches leveraging all available information provides an ideal data ecosystem to support dose selection/optimization and assess risk/benefit profile (**Figure 3**).

Although GT has the potential to treat or even cure rare genetic diseases, unexpected cellular and/or humoral immune response to the vector and/or transgene can constitute significant hurdles in clinical development of AAV based GT.[92] Modeling approaches, such as QSP, to predict innate and/or adaptive immune response to GT and any resultant tissue damage will be of tremendous value in *a priori* assessment of risk/benefit. Currently available models[93] could be a good starting point, however, given the complex mechanism of action of AAV-based GT and lack of thorough understanding of its pharmacology, it is of paramount importance that the knowledge gaps and assumptions in modeling GTs are clearly understood and explained. Even after significant advances in the field, there are several gaps in knowledge that limit the development of models, especially mechanistic approaches such as QSP and PBPK: (i) vector uptake and trafficking, (ii) development of immunogenicity, (iii) measuring gene expression in animals and humans, and (iv) effect of empty vectors and batch-to-batch variability in manufacture. The learn-and-confirm approach to continuously develop and improve the models based on emerging data is needed.

A recently published landscape analysis from the FDA,[94] shows a significant increase in use of artificial intelligence (AI) and ML in regulatory submissions in 2021 (132 AI/ML components) as compared with 14, 7, 3, 1, and 1 in 2020, 2019, 2018, 2017, and 2016, respectively. Specifically in GT, AI and ML approaches can facilitate—(i) accurate identification of the target gene to increase probability of therapeutic success and reduce off-target effects, (ii) exploration of vast design space for optimizing the design of transgene and vector, and finally (iii) designing of clinical studies by identifying the right patient population, optimal PK and PD sampling, and predicting therapeutic and adverse event profiles. However, several challenges need to be overcome to fully realize the value of these approaches, such as adequate training and validation of AI and ML systems using independent datasets, due to limited publicly available experimental data, and adoption of Good Machine Learning Practice in drug development.[95]

In this paper, challenges and complexities in conducting GT trials have been highlighted. These scientific issues are compounded by ethical considerations, such as fairness in patient selection, and appropriate patient education to understand the potential risks (and benefits) of GT. We have discussed several strategies to implement MIDD in conjunction with other approaches and sources of data to improve GT clinical trials. The use of validated surrogate end points is key in utilizing MIDD in rare diseases drug development. One such example is the multi-luminance mobility test (MLMT) in inherited retinal diseases,[62] which has been used for approval of voretigene

Figure 3 The model-informed drug development ecosystem for development of adeno-associated virus vector-based gene therapy, enabled using novel clinical strategies.

neparvovec-rzyl.[96] The application of modeling approaches by incorporating end points like MLMT to support dose selection should be considered. Quantitative approaches to model NH data from prospective multicenter trials or from published case reports is crucial to increase confidence in these data, as such data could have been generated in a very small number of patients or without harmonized protocol. Garbade *et al*.,[97] showed the utility of several statistical methodologies of modeling NH data in seven ultra-rare neurogenetic diseases. Similarly, the use of a quantitative biomarker-based approach to track NH and application in trial enrichment was demonstrated in patients with rare autosomal dominant polycystic kidney disease (ADPKD).[98] Perrone *et al*.[98] developed a statistical model to link longitudinal total kidney volume (an imaging biomarker), age and estimated glomerular filtration rate (eGFR) to the probability of a 30% decline of eGFR or end-stage renal disease and ultimately leveraged the model for trial enrichment in patients with ADPKD. Such approaches should be adopted for AAV-GT trials to improve confidence in NH data. Modeling of maximum or minimum treatment effects of GT, using the concepts of extreme value theory,[65] to predict rare safety events, such as vision loss, could provide more context and confidence around long-term clinical effects. An additional approach to streamline GT development, especially in rare diseases is to conduct platform trials using a durable master protocol.[99] Although significant strides have been made to make platform trials a reality or coronavirus disease 2019 (COVID-19) drugs and vaccines,[100] legal and regulatory hurdles must be overcome to make them a reality in GT.

In closing, although GT products have profound transformative potential after a single dose, they also carry unique and complex risks. Thus, a combination of MIDD approaches and innovative clinical trial design is required to provide greater insights into the safety and efficacy of GT products, ultimately benefiting the patients.

FUNDING

No funding was received for this work.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

FINANCIAL DISCLOSURES

A.M. is an employee of Kura Oncology and holds shares of Kura oncology. M.A.A., K.S., F.D.G., and O.C. are employees of Takeda. R.K. is an employee of Certara. N.R. is an employee of Parexel. Y.R. is currently employed at the US FDA. S.A. is an employee of King Abdulaziz University. M.R. is an employee of UCB Biopharma. I.Y. is an employee of Gilead Science.

1. Bulaklak, K. & Gersbach, C.A. The once and future gene therapy. *Nat. Comm.* **11**, 1–4 (2020).
2. Dunbar, C.E., High, K.A., Joung, J.K., Kohn, D.B., Ozawa, K. & Sadelain, M. Gene therapy comes of age. *Science* **359**, 6372 (2018).
3. Wang, D., Tai, P.W.L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Dis.* **18**, 358–378 (2019).
4. Qosa, H., Hassan, H.E. & Younis, I.R. Overview of clinical pharmacology packages of new drug applications approved for the treatment of rare diseases. *J. Clin. Pharmacol.* **62**, S72–S78 (2022).
5. US FDA. Cellular & gene therapy guidelines. <https://www.fda.gov/vaccines-blood-biologics/biologics-guidances/cellular-gene-therapy-guidances>. Accessed February 15, 2023.
6. European Medicines Agency. Multidisciplinary gene therapy guidelines. <https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-guidelines/multidisciplinary/multidisciplinary-gene-therapy>. Accessed February 15, 2023.
7. Sun, K. & Liao, M.Z. Clinical pharmacology considerations on recombinant adeno-associated virus-based gene therapy. *J. Clin. Pharmacol.* **62**, S79–S94 (2022).
8. Chen, N., Sun, K., Chemuturi, N.V., Cho, H. & Xia, C.Q. The perspective of DMPK on recombinant adeno-associated virus-based gene therapy: past learning, current support, and future contribution. *AAPS J.* **24**, 31–45 (2022).
9. Tang, F., Wong, H. & Chee, M.N. Rational clinical dose selection of adeno-associated virus-mediated gene therapy based on allometric principles. *Clin. Pharmacol. Ther.* **110**, 803–807 (2021).
10. Aksenov, S., Roberts, J.C., Mugundu, G., Mueller, K.T., Bhattacharya, I. & Tortorici, M.A. Current and next steps toward prediction of human dose for gene therapy using translational dose-response studies. *Clin. Pharmacol. Ther.* **110**, 1176–1179 (2021).
11. Belov, A., Schultz, K., Forshee, R. & Tegenge, M.A. Opportunities and challenges for applying model-informed drug development approaches to gene therapies. *CPT Pharmacometrics Syst. Pharmacol.* **10**, 286–290 (2021).
12. Li, R.J. *et al*. Model-informed approach supporting drug development and regulatory evaluation for rare diseases. *J. Clin. Pharmacol.* **62**, S27–S37 (2022).
13. Mitra, A. & Wang, Y. Applications of model informed drug development (MIDD) in drug development lifecycle and regulatory review. *Pharm. Res.* **39**, 1663–1667 (2022).
14. US FDA. Human Gene Therapy for Rare Diseases. Guidance for industry.
15. Zou, P. Interspecies normalization of dose-response relationship for adeno-associated virus-mediated haemophilia gene therapy—application to human dose prediction. *Br. J. Clin. Pharmacol.* **89**, 1393–1401 (2022).
16. Zou, P. First-in-patient dose prediction for adeno-associated virus-mediated hemophilia gene therapy using allometric scaling. *Mol. Pharm.* **20**, 758–766 (2023).
17. Zolgensma® prescribing information. <https://www.fda.gov/media/126109/download>. Accessed February 15, 2023.
18. Roctavian® summary of product characteristics. <https://www.ema.europa.eu/en/documents/product-information/roctavian-epar-product-information_en.pdf>. Accessed February 15, 2023.
19. Hemgenix® prescribing information. <https://labeling.cslbehring.com/PI/US/Hemgenix/EN/Hemgenix-Prescribing-Information.pdf>. Accessed February 15, 2023.
20. Evers, M. *et al*. Translating preclinical data to a human equivalent dose for AMT-130 AAV gene therapy for early manifest huntington's disease. *Mov. Disord.* **34**, S7–S8 (2019).
21. Upstaza® EMA assessment report. EMA/CHMP/571076/2022. <https://www.ema.europa.eu/en/documents/assessment-report/upstaza-epar-public-assessment-report_en.pdf>. Accessed February 15, 2023.
22. Burr, A., Erickson, P., Bento, R., Shama, K., Roth, C. & Parekkadan, B. Allometric-like scaling of AAV gene therapy for systemic protein delivery. *Mol. Ther. Methods Clin. Dev.* **27**, 368–379 (2022).
23. West, G.B., Brown, J.H. & Enquist, B.J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126 (1997).
24. Hatton, I.A., Dobson, A.P., Storch, D. & Loreau, M. Linking scaling laws across eukaryotes. *Proc. Natl. Acad. Sci.* **116**, 21616–21622 (2019).
25. Thomsen, G. *et al*. Biodistribution of onasemnogene abeparvovec DNA, mRNA and SMN protein in human tissue. *Nat. Med.* **27**, 1701–1711 (2021).
26. Karumuthil-Melethil, S. *et al*. Gene expression from AAV vectors in the liver: a comparative study across species, promoters and AAV serotypes (abstract 824). *Mol. Ther* **29**(Supplement 1), 389 (2021).
27. Nietupski, J.B. *et al*. Systemic administration of AAV8-α-galactosidase a induces humoral tolerance in nonhuman primates despite low hepatic expression. *Mol. Ther.* **19**, 1999–2011 (2011).
28. Fong, S. *et al*. Interindividual variability in transgene mRNA and protein production following adeno-associated virus gene therapy for hemophilia a. *Nat. Med.* **28**, 789–797 (2022).
29. Bradshaw, E.L. *et al*. Applications of quantitative systems pharmacology in model-informed drug discovery: perspective on impact and opportunities. *CPT Pharmacometrics Syst. Pharmacol.* **8**, 777–791 (2019).
30. Jeon, J.Y., Ayyar, V.S. & Mitra, A. Pharmacokinetic and pharmacodynamic modeling of siRNA therapeutics—a minireview. *Pharm. Res.* **39**, 1749–1759 (2022).
31. Zheng, B. *et al*. A systems pharmacology model for gene therapy in sickle cell disease. *CPT Pharmacometrics Syst. Pharmacol.* **10**, 696–708 (2021).
32. Sun, K.F., Zhang, Z.W., Ko, G. & Bradshaw, E.L. Physiologically based pharmacokinetic modeling for the biodistribution of adeno-associated virus serotype 8 after intravenous Administration in Mice and non-Human Primates. *Mol. Ther.* **29**, 129–130 (2021).
33. Rao, S., *et al*. Developing a robust Quantitative Systems Pharmacology model of adeno-associated virus (AAV) based gene therapy for clinical applications. Presented at the 12th American Conference for Pharmacometrics. <https://www.eposters.net/poster/developing-a-robust-quantitative-systems-pharmacology-model-of-adeno-associated-virus-aav-based>. Published 2021. Accessed February 15, 2023.
34. Shah, D. & Betts, A.M. Towards a platform PBPK model to characterize the plasma and tissue disposition of monoclonal antibodies in preclinical species and human. *J. Pharmacokinet. Pharmacodyn.* **39**, 67–86 (2012).
35. Gill, K., Gardner, I., Li, L. & Jamei, M. A bottom-up whole-body physiologically based pharmacokinetic model to mechanistically predict tissue distribution and the rate of subcutaneous absorption of therapeutic proteins. *AAPS J.* **18**, 156–170 (2016).
36. Chen, X., Hickling, T.P. & Vicini, P. A mechanistic, multiscale mathematical model of immunogenicity for therapeutic proteins: part 1—theoretical model. *CPT Pharmacometrics Syst. Pharmacol.* **3**, e133 (2014).
37. Chen, X., Hickling, T.P. & Vicini, P. A mechanistic, multiscale mathematical model of immunogenicity for therapeutic proteins: part 2—model applications. *CPT Pharmacometrics Syst. Pharmacol.* **3**, e134 (2014).
38. Schwanhausser, B. *et al*. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
39. US FDA Summary Basis for Regulatory Action. onasemnogene abeparvovec-xioi. <https://www.fda.gov/media/127961/download>. Accessed March 09, 2023.
40. Meyer, K. *et al*. Improving single injection CSF delivery of AAV9-mediated gene therapy for SMA: a dose-response study in mice and nonhuman primates. *Mol. Ther.* **23**, 477–487 (2015).
41. Beachy, S.H., Alper, J., Hackmann, M., & Addie, S. Exploring novel clinical trial designs for gene-based therapies: Proceedings of a workshop. https://doi.org/10.17226/25712 (2020).
42. Thorlund, K., Dron, L., Park, J.J.H. & Mills, E.J. Synthetic and external controls in clinical trials—a primer for researchers. *Clin. Epidemiol.* **12**, 457–467 (2020).

43. US FDA. Considerations for the design and conduct of externally controlled trials for drug and biological products. Guidance for industry. February 2023. <https://www.fda.gov/media/164960/download>. Accessed March 09, 2023.
44. US FDA Summary Basis for Regulatory Action. Blinatumomab. <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2014/125557Orig1s000TOC.cfm>. Accessed May 20, 2023.
45. Konkle, B. *et al*. Core data set on safety, efficacy, and durability of hemophilia gene therapy for a global registry: communication from the SSC of the ISTH. *J. Thromb. Haemost.* **18**, 3074–3077 (2020).
46. Garrison, L.P. *et al*. Hemophilia gene therapy value assessment: methodological challenges and recommendations. *Value Health* **24**, 1628–1633 (2021).
47. Liu, J. *et al*. Natural history and real-world data in rare diseases: applications, limitations, and future perspectives. *J. Clin. Pharmacol.* **62**, S38–S55 (2022).
48. US FDA. Rare diseases: natural history studies for drug development. Guidance for Industry. <https://www.fda.gov/media/122425/download>. Accessed February 15, 2023.
49. Cances, C. *et al*. Natural history of type 1 spinal muscular atrophy: a retrospective, global, multicenter study. *Orphanet J. Rare Dis.* **17**, 300 (2022).
50. Chung, D.C. *et al*. The natural history of inherited retinal dystrophy due to biallelic mutations in the RPE65 gene. *Am. J. Ophthalmol.* **199**, 58–70 (2019).
51. Newman, N.J. *et al*. Intravitreal gene therapy vs. natural history in patients with Leber hereditary optic neuropathy carrying the m.11778G>a *ND4* mutation: systematic review and indirect comparison. *Front. Neurol.* **12**, 662838 (2021).
52. Tee, J.J.L., Yang, Y., Kalitzeos, A., Webster, A., Bainbridge, J. & Michaelides, M. Natural history study of retinal structure, progression, and symmetry using ellipzoid zone metrics in RPGR-associated retinopathy. *Am. J. Ophthalmol.* **198**, 111–123 (2019).
53. Nguyen, X.T.A. *et al*. Clinical characteristics and natural history of rho-associated retinitis pigmentosa: a long term follow-up study. *Retina* **41**, 213–223 (2021).
54. Brogna, C. *et al*. Long-term natural history data in Duchenne muscular dystrophy ambulant patients with mutations amenable to skip exons 44, 45, 51 and 53. *Plos One* **14**, e0218683 (2019).
55. Nickel, M. & Schulz, A. Natural history studies in NCL and their expanding role in drug development: experiences from CLN2 disease and relevance for clinical trials. *Front. Neurol.* **13**, 785841 (2022).
56. Day, J.W. *et al*. Onasemnogene abeparvovec gene therapy for symptomatic infantile-onset spinal muscular atrophy in patients with two copies of *SMN2* (STR1VE): an open-label, single-arm, multicentre, phase 3 trial. *Lancet Neurol.* **20**, 284–293 (2021).
57. Mercuri, E. *et al*. Onasemnogene abeparvovec gene therapy for symptomatic infantile-onset spinal muscular atrophy type 1 (STR1VE-EU): an open-label, single-arm, multicentre, phase 3 trial. *Lancet Neurol.* **20**, 832–841 (2021).
58. Carvalho, M., Martins, A.P. & Sepodes, B. Hurdles in gene therapy regulatory approval: a retrospective analysis of European marketing authorization applications. *Drug Discov. Today* **24**, 823–828 (2019).
59. Ozelo, M.C. *et al*. Valoctocogene Roxaparvovec gene therapy for hemophilia a. *N. Engl. J. Med.* **386**, 1013–1025 (2022).
60. Roctavian® EMA assessment report. EMA/685615/2022. <https://www.ema.europa.eu/en/documents/assessment-report/roctavian-epar-public-assessment-report_en.pdf>. Accessed February 15, 2023.
61. Hemgenix® BLA clinical review memorandum. <https://www.fda.gov/media/164332/download>. Accessed February 15, 2023.
62. Chung, D. *et al*. Novel mobility test to assess functional vision in patients with inherited retinal dystrophies. *Clin. Experiment. Ophthalmol.* **46**, 247–259 (2018).
63. McIntosh, A., Sverdlov, O., Yu, L. & Kaufmann, P. Clinical design and analysis strategies for the development of gene therapies: considerations for quantitative drug development in the age of genetic medicine. *Clin. Pharmacol. Ther.* **110**, 1207–1215 (2021).
64. Walter, E. & Pronzato, L. Optimal experiment design for nonlinear models subject to large prior uncertainties. *Am. J. Physiol.* **253**, R530–R534 (1987).
65. Bonate, P.L. The application of extreme value theory to pharmacometrics. *J. Pharmacokinet. Pharmacodyn.* **48**, 83–97 (2021).
66. Chan, P., Peskov, K. & Song, X. Applications of model-based meta-analysis in drug development. *Pharm. Res.* **39**, 1761–1777 (2022).
67. Lennie, J.L., Mondick, J.T. & Gastonguay, M.R. Bayesian modeling and simulation to inform rare disease drug development early decision-making: application to Duchenne muscular dystrophy. *PLoS One* **17**, e0247286 (2022).
68. Shah, J., Kim, H., Sivamurthy, K., Monahan, P.E. & Fries, M. Comprehensive analysis and prediction of long-term durability of factor IX activity following etranacogene dezaparvovec gene therapy in the treatment of hemophilia B. *Curr. Med. Res. Opin.* **39**, 227–237 (2023).
69. Nathwani, A.C. *et al*. Adeno-associated mediated gene transfer for hemophilia B: 8 year follow up and impact of removing "empty viral particles" on safety and efficacy of gene transfer. *Blood* **132**(Suppl 1), 491 (2018).
70. Cook, K., Forbes, S.P., Adamski, K., Ma, J.J., Chawla, A. & Garrison, L.P. Assessing the potential cost-effectiveness of a gene therapy for the treatment of hemophilia a. *J. Med. Econ.* **23**, 501–512 (2020).
71. O'Hara, J. & Neumann, P.J. Health technology assessment for gene therapies in haemophilia. *Haemophilia* **28**(Suppl 2), 19–26 (2022).
72. Krishna, R. The utility of model-informed drug development for rare diseases. <https://www.certara.com/blog/the-utility-of-model-informed-drug-development-for-rare-diseases/>. Accessed February 15, 2023.
73. Temple, R. Enrichment of clinical study populations. *Clin. Pharmacol. Ther.* **88**, 774–778 (2010).
74. Green, D.J. *et al*. Enrichment strategies in pediatric drug development: an analysis of trials submitted to the US Food and Drug Administration. *Clin. Pharmacol. Ther.* **104**, 983–988 (2018).
75. US FDA. Enrichment strategies for clinical trials to support determination of effectiveness of human drugs and biological products. Guidance for Industry. <https://www.fda.gov/media/121320/download>. Accessed February 15, 2023.
76. Bennett, J. *et al*. Safety and durability of effect of contralateral-eye administration of AAV2 gene therapy in patients with childhood-onset blindness caused by RPE65 mutations: a follow-on phase 1 trial. *Lancet* **388**, 661–672 (2016).
77. Russell, S. *et al*. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet* **390**, 849–860 (2017).
78. Madabushi, R., Seo, P., Zhao, L., Tegenge, M. & Zhu, H. Review: role of model-informed drug development approaches in the lifecycle of drug development and regulatory decision-making. *Pharm. Res.* **39**, 1669–1680 (2022).
79. Haber, G. *et al*. Association of genetic mutations and loss of ambulation in childhood-onset dystrophinopathy. *Muscle Nerve* **63**, 181–191 (2021).
80. Ronney, W.D. *et al*. Modeling disease trajectory in Duchenne muscular dystrophy. *Neurology* **94**, e1622–e1633 (2020).
81. US FDA. Framework for FDA's real-world evidence program. <https://www.fda.gov/media/120060/download>. Accessed February 15, 2023.
82. Cave, A., Kurz, X. & Arlett, P. Real-world data for regulatory decision making: challenges and possible solutions for Europe. *Clin. Pharmacol. Ther.* **106**, 36–39 (2019).

83. Sherman, R.E. *et al*. Real-world evidence—what is it and what can it tell us? *N. Engl. J. Med.* **375**, 2293–2297 (2016).
84. Long-term follow-up study in subjects who received voretigene neparvovec-rzyl (AAV2-hRPE65v2). <https://clinicaltrials.gov/ct2/show/NCT03602820>. Accessed February 15, 2023.
85. Fischer, M.D., Maier, R., Suhner, A., Stiehl, D., Fasser, C. & Leroy, B.P. PERCEIVE study report: real world safety and effectiveness of voretigene neparvovec. *IOVS* **63**, 451 (2022).
86. Lee, S. *et al*. Short-term clinical outcomes of onasemnogene abeparvovec treatment for spinal muscular atrophy. *Brain Dev.* **44**, 287–293 (2022).
87. Press release—Data from health care should close gaps in the evidence for new medicines. <https://www.g-ba.de/presse/pressemitteilungen-meldungen/932/>. Accessed February 15, 2023.
88. Darrow, J.J. Luxturna: FDA documents reveal the value of a costly gene therapy. *Drug Discov. Today* **24**, 949–954 (2019).
89. Qiu, T. *et al*. Gene therapy evidence generation and economic analysis: pragmatic considerations to facilitate fit-for-purpose health technology assessment. *Front. Public Health* **10**, 773629 (2022).
90. Iglesias-Lopez, C. *et al*. Current andscape of clinical development and approval of advanced therapies. *Mol. Ther. Methods Clin. Dev.* **23**, 606–618 (2021).
91. Nukala, U. *et al*. A systematic review of the efforts and hindrances of modeling and simulation of CAR T-cell therapy. *AAPS J.* **23**, 52 (2021).
92. Shirley, J.L. *et al*. Immune responses to viral gene therapy vectors. *Mol. Ther.* **28**, 709–722 (2020).
93. Kierzek, A.M. *et al*. A quantitative systems pharmacology consortium approach to managing immunogenicity of therapeutic proteins. *CPT Pharmacometrics Syst. Pharmacol.* **8**, 773–776 (2019).
94. Liu, Q. *et al*. Landscape analysis of the application of artificial intelligence and machine learning in regulatory submissions for drug development from 2016 to 2021. *Clin. Pharmacol. Ther.* **113**, 771–774 (2022).
95. US FDA. Good machine learning practice for medical device development: Guiding principles. <https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles>. Accessed March 10, 2023.
96. Luxturna® prescribing information. <https://www.fda.gov/media/109906/download>. Accessed March 10, 2023.
97. Garbade, S.F. *et al*. Quantitative retrospective natural history modeling for orphan drug development. *J. Inherit. Metab. Dis.* **44**, 99–109 (2021).
98. Perrone, R.D. *et al*. A drug development tool for trial enrichment in patients with autosomal dominant polycystic kidney disease. *Kidney Int. Rep.* **2**, 451–460 (2017).
99. Brooks, P.J. *et al*. The platform vector gene therapies project: increasing the efficiency of adeno-associated virus gene therapy clinical trial startup. *Hum. Gene Ther.* **31**, 1034–1042 (2020).
100. Vanderbeek, A.M., Bliss, J.M., Yin, Z. & Yap, C. Implementation of platform trials in the COVID-19 pandemic: a rapid review. *Contemp. Clin. Trials* **112**, 106625 (2022).